CORRELATION BETWEEN A PARAMETER AND AN ESTIMATE: FACTS, FORMULAE, EXAMPLES

by

Anirban DasGupta and William Strawderman
Purdue University | Rutgers University

Technical Report #96-30

Department of Statistics
Purdue University
West Lafayette, IN USA

July 1996

CORRELATION BETWEEN A PARAMETER AND AN ESTIMATE: FACTS, FORMULAE, EXAMPLES*

by

Anirban DasGupta and William Strawderman
Purdue University Rutgers University

Abstract

We consider correlations between a parametric function and an estimate as an index of goodness of the proposed estimate. The results include a general formula, which specializes to neat forms for unbiased estimates and Bayes estimates induced by the joint probability space. The correlation criterion is shown to have connections to other criteria and various positive properties of unbiased and Bayesian estimates are established by consideration of their correlation with a parameter. Also presented are a nonparametric Bayes risk identity with an application to confidence set estimation in Exponential families and several examples that illustrate a number of interesting phenomena related to the topic.

1. Introduction

In this article we present an intuitively interesting criterion of goodness of an estimate of a parameter: correlation between the parameter and the estimate. Naturally, this makes sense only in a formally Bayesian framework. That is, one must have a joint probability space of the observable and the parameter and furthermore the estimate and the parameter each must be a L_2 function in the joint probability space. Thus, at the very starting point, one has, in addition to a sampling model a prior distribution on the parameter as well. One may or may not want to think of it as a subjectively elicited prior, depending on taste. Actually, the correlation between the parameter and any linear function of a given estimate is the same as the correlation between the parameter and the estimate itself; thus, correlation is a possible criterion to assess only equivalence classes of estimates, all affinely equivalent estimates being treated as the same. In a given context, one member of an equivalence class should be the natural estimate. For example, for estimating a location

* Research supported by NSF grants DMS 93–07727 and 94–00476.

parameter, the median seems a better choice than twice the median.

If one thinks of Bayes estimates with respect to squared error loss as the regression of the parameter on the observation, then consideration of correlation comes in naturally; after all, the square of the correlation is an index of how useful the regression was. Going one step further, one then looks at correlations between the parameter and a general estimate, not merely the Bayes estimate with respect to the specified prior. This seems even more natural when many different priors are considered feasible and one wants to look for estimates that are highly correlated with the parameter under all or many of these feasible priors.

As our calculations will clearly show, correlation is intimately connected to other established criteria of statistical theory. In particular, our calculations will show its connection to the minimax and the Bayes criteria. A nice thing about considering correlations is that we see these connections to other accepted criteria, and in addition it is necessarily between -1 and 1. Indeed, for reasonable estimates, it will be between 0 and 1. Thus it is more easily interpretable, unlike a full risk function or a Bayes risk, which have no theoretical universal bounds in general.

In section 2, we give a general correlation formula and exhibit that it simplifies to two pretty formulae for unbiased estimates and Bayes estimates. In the process, we give a nonparametric Bayes risk identity connecting the Bayes risk to the bias of the Bayes estimate. An old theorem about the impossibility of Bayes estimates being unbiased follows immediately from this identity.

In section 3, we establish a variety of connections of the correlation criterion to other criteria. We show, in particular, that all unbiased estimates are always positively correlated with the parameter. Bayes estimates have a slightly weaker property. We give a general result on the multiparameter Exponential family, and we show that in general compact parameter spaces, the Bickel-Levit prior has a certain minimum correlation property. We also show that a UMVUE is always more correlated with the parameter than any other unbiased estimate and give a general result on automatic construction of a posterior credible set in the multiparameter Exponential family by making use of the nonparametric Bayes risk identity in section 2.

Section 4 gives illustrative examples of a number of different phenomena.

The principal contributions of this article are the following:

a. We present an intuitively interesting index of goodness of an estimate and present some apparently novel calculations;

b. We establish connections to other criteria;

c. We show positive properties of unbiased and other types of estimates vis-a-vis this criterion;

d. We show a general correlation minimizing property of the Bickel-Levit prior;

e. We give a number of illustrative examples that show various interesting phenomena;

f. We give a nonparametric Bayes risk identity and an application to automatic construction of posterior credible sets in the general multiparameter Exponential family.

2. Four Identities

In this section, we will first derive a general correlation formula between a parameter and an estimate in a general probability space $L_2(\mathcal{X} \otimes (H), P)$. We will apply the formula to unbiased estimates and the Bayes estimate corresponding to the prior distribution induced by the probability measure P on the joint space. It will be seen that the correlation formula takes a nice form for each of these two cases.

Proposition 1. Let $\delta(X)$ be any estimate of θ; then the correlation between θ and δ in the joint probability space $L_2(\mathcal{X} \otimes (H), P)$ equals

$$\rho_\pi(\theta, \delta) = \frac{V(\pi) + \operatorname{Cov}_\pi(\theta, b(\theta))}{\sqrt{V(\pi)\{r(\pi,\delta) + \operatorname{var}_\pi(\theta + b(\theta)) - E_\pi(b^2(\theta))\}}} \tag{2.1}$$

where π denotes the prior distribution induced by P, $V(\pi)$ denotes the variance of θ under π, $b(\theta) = E_{X|\theta}\delta(X) - \theta$ denotes the bias of the estimate $\delta(X)$, Cov denotes covariance under π, $E_\pi(\cdot)$ denotes expectation with respect to π, and $r(\pi, \delta) = E_\pi E_{X|\theta}(\delta(X) - \theta)^2$ denotes the Bayes risk of δ under π.

Proof: We can assume without loss of generality that $E_\pi(\theta) = 0$. Then, the covariance of θ and δ is

$$\begin{aligned}\text{Cov }(\theta,\delta) = E(\theta\delta(X)) = E_\pi(\theta\cdot E_{X|\theta}\delta(X)) &= E_\pi(\theta^2+\theta b(\theta)) \\ &= V(\pi) + \text{ Cov }(\theta, b(\theta)).\end{aligned} \tag{2.2}$$

Next, the marginal variance of δ is

$$\begin{aligned}\text{var}_P(\delta(X)) &= E_\pi \text{ var}_{X|\theta}\delta(X) + \text{ var}_\pi E_{X|\theta}\delta(X) \\ &= E_\pi\{E_{X|\theta}(\delta(X)-\theta)^2 - b^2(\theta)\} + \text{ var}_\pi(\theta+b(\theta)) \\ &= r(\pi,\delta) + \text{ var}_\pi(\theta+b(\theta)) - E_\pi(b^2(\theta)).\end{aligned} \tag{2.3}$$

The proposition is therefore proved. □

Corollary 1. Let $\delta(X)$ be any unbiased estimate of θ. Then,

$$\rho_\pi(\theta,\delta) = \sqrt{\frac{V(\pi)}{V(\pi)+r(\pi,\delta)}}. \tag{2.4}$$

Corollary 1 follows immediately from (2.1) as the bias $b(\theta)\equiv 0$. □

To derive a similar nice formula when $\delta(X)$ equals the Bayes estimate $\delta_\pi(X) = E_P(\theta|X)$ itself, we first need the following nonparametric Bayes risk identity. We find the identity quite intriguing in the way it connects Bayes risk to the bias of the Bayes estimate.

Proposition 2. Let $b_\pi(\theta) = E_{X|\theta}\delta_\pi(X) - \theta$ denote the bias of the Bayes estimate $\delta_\pi(X)$. Then the Bayes risk $r(\pi) = E_\pi E_{X|\theta}(\delta_\pi(X)-\theta)^2$ equals

$$r(\pi) = -E_\pi(\theta b_\pi(\theta)). \tag{2.5}$$

The multiparameter version of (2.5) is

$$r(\pi) = -E_\pi(\underset{\sim}{\theta}'\underset{\sim}{b}_\pi(\underset{\sim}{\theta})) \tag{2.5'}$$

First we point out the following well known fact, which follows as a corollary of our Proposition 2:

Corollary 2. The Bayes estimate $\delta_\pi(X)$ cannot be unbiased unless $r(\pi) = 0$.

Other proofs of Corollary 2 can be seen in Ferguson (1973) and Bickel and Blackwell (1967). We also point out the following new consequence of Proposition 2 before we get back to our discussion of correlations.

Corollary 3. Let $\nu = E_\pi|\theta|$ and $M = \sup_\theta |b_\pi(\theta)|$. Then

$$r(\pi) \leq \nu M. \tag{2.6}$$

(2.6) also holds in the multiparameter case with $\nu = E_\pi||\underset{\sim}{\theta}||_2$ and $M = \sup_{\underset{\sim}{\theta}} ||\underset{\sim}{b}_\pi(\underset{\sim}{\theta})||_2$. Thus, it is useful to use such Bayes estimates whose bias stay small; we find (2.6) quite interesting. Now we will give a proof of Proposition 2 itself.

Proof of Proposition 2: First note that

$$\begin{aligned}
r(\pi) &= E_P(\delta_\pi(X) - \theta)^2 \\
&= E_\pi E_{X|\theta}(\delta_\pi(X) - \theta)^2 \\
&= E_\pi E_{X|\theta}(\delta_\pi^2(X) + \theta^2 - 2\theta\delta_\pi(X)) \\
&= E_P\delta_\pi^2(X) + E_\pi\theta^2 - 2E_\pi\theta(\theta + b_\pi(\theta)) \\
&= E_P\delta_\pi^2 - E_\pi\theta^2 - 2E_\pi(\theta b_\pi(\theta)).
\end{aligned} \tag{2.7}$$

On the other hand,

$$\begin{aligned}
r(\pi) &= E_P(\delta_\pi(X) - \theta)^2 \\
&= E_P E_{\theta|X}(\delta_\pi(X) - \theta)^2 \\
&= E_P E_{\theta|X}(\theta^2 - \delta_\pi^2(X)) \\
&= E_\pi(\theta^2) - E_P\delta_\pi^2(X).
\end{aligned} \tag{2.8}$$

Therefore, from (2.7), we have

$$\begin{aligned}
r(\pi) &= -r(\pi) - 2E_\pi(\theta b_\pi(\theta)) \\
\Rightarrow r(\pi) &= -E_\pi(\theta b_\pi(\theta)), \text{ Q.E.D.}
\end{aligned}$$

□

Use of Proposition 1 and Proposition 2 results in a neat formula for the correlation between θ and δ_π, as is shown below.

Corollary 4.

$$\rho_\pi(\theta, \delta_\pi) = \sqrt{1 - \frac{r(\pi)}{V(\pi)}}. \tag{2.9}$$

Proof: We will use (2.1). The numerator

$$\begin{aligned}
&V(\pi) + \operatorname{Cov}_\pi(\theta, b_\pi(\theta)) \\
&= V(\pi) + E_\pi(\theta b_\pi(\theta)) - E_\pi(\theta)E_\pi(b_\pi(\theta)) \\
&= V(\pi) + E_\pi(\theta b_\pi(\theta)) \qquad (\text{ as } E_\pi(b_\pi(\theta)) = 0) \\
&= V(\pi) - r(\pi). \qquad (2.10) \\
&\qquad \text{(by Proposition 2)}
\end{aligned}$$

Also,

$$\begin{aligned}
&r(\pi, \delta_\pi) + \operatorname{var}_\pi(\theta + b_\pi(\theta)) - E_\pi(b_\pi^2(\theta)) \\
&= r(\pi) + V(\pi) + E_\pi b_\pi^2(\theta) + 2E_\pi(\theta b_\pi(\theta)) - E_\pi(b_\pi^2(\theta)) \\
&= V(\pi) - r(\pi). \qquad (2.11) \\
&\qquad \text{(again by Proposition 2)}
\end{aligned}$$

$$\begin{aligned}
\therefore \quad \rho_\pi(\theta, \delta_\pi) &= \frac{V(\pi) - r(\pi)}{\sqrt{V(\pi)(V(\pi) - r(\pi))}} \\
&= \sqrt{1 - \frac{r(\pi)}{V(\pi)}},
\end{aligned}$$

as stated. □

3. Further Consequences

We now use our correlation formulae of section 2 to obtain a number of connections to other established methods and concepts of statistics. These will all be gathered in a single proposition stated below.

Proposition 3.

a. The correlation between θ and any unbiased estimate is always nonnegative and strictly positive unless the prior is degenerate.

b. If θ has a UMVUE δ_0, then $\rho_\pi(\theta, \delta_0) \geq \rho_\pi(\theta, \delta_u)$ for any other unbiased estimate δ_u and the inequality is strict if δ_0 is the unique UMVUE and π gives support to the entire parameter space.

c. $\rho_\pi(\theta, \delta_\pi)$ is also always nonnegative (but can be zero even for nondegenerate priors).

d. In the general Exponential family, an estimate $\delta(X)$ of the mean $\mu = E(X)$ always has nonnegative correlation with μ if δ is monotone nondecreasing in the observation X.

e. If a least favorable prior with a finite variance exists, say π_0, then π_0 always minimizes $\rho_\pi(\theta, \delta_\pi)$ among all π with $V(\pi) \geq V(\pi_0)$.

f. If the parameter space is a compact interval $[-m, m]$, then up to the order $o(n^{-2})$, the Bickel-Levit prior π_m minimizes $\rho_\pi(\theta, \delta_\pi)$ among all π with $V(\pi) \geq V(\pi_m) = \frac{m^6\pi^3 + 6m^2\pi\cos(m^2\pi) - 6\sin(m^2\pi) + 3m^4\pi^2\sin(m^2\pi)}{3m^2\pi^2(m^2\pi + \sin(m^2\pi))}$ (e.g., $V(\pi_1) = .1307,\ V(\pi_5) = 8.3252$).

g. In a location parameter model, i.e., if $X \sim F(x-\theta)$ and $\theta \sim \pi$, and F, π belong respectively to specified classes $\mathcal{C}_1, \mathcal{C}_2$, the criterion of maximizing the minimum correlation $\inf_{F,\pi} \rho_{F,\pi}(\theta, \delta)$ over all estimates δ which are unbiased under each F is equivalent to the common minimax criterion of minimizing $\sup_{\mathcal{C}_1} \operatorname{var}_F(\delta)$.

Discussion of Proposition 3. a is a nice property of arbitrary unbiased estimates and b of the UMVUE. c is expected, but that it can be zero in nontrivial cases surprised us. We shall see an example in section 4. d is a nice general result in the Exponential family. e says that if we construct a noninformative prior by minimizing $\rho_\pi(\theta, \delta_\pi)$, we will get the least favorable prior π_0 if other candidate priors have a variance as large. For example, take the Binomial problem: the Beta $(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$ prior is least favorable and has variance $\frac{1}{4(1+\sqrt{n})} \leq .05$ if $n \geq 16$. So the Beta $(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$ prior is obtained as a noninformative prior under the very mild restriction $V(\pi) \geq .05$ if $n \geq 16$. f is an interesting general property of the Bickel-Levit prior, *not* isolated to the bounded normal mean case. g gives another connection of correlation between a parameter and an estimate to other established optimality criteria of statistics.

Proof of Proposition 3: Both a and b are transparent from (2.4) and c follows from (2.9) as $r(\pi) \leq V(\pi)$. To see (d), observe that for any estimate $\delta(X)$, $\text{cov}_\pi(\theta, \delta) = E_P(\theta\delta(X)) - E_\pi(\theta)E_P(\delta(X)) = E_P(\delta_\pi(X)\delta(X)) - E_P(\delta_\pi(X))E_P(\delta(X)) = \text{cov}_P(\delta_\pi, \delta) \geq 0$ if δ is monotone nondecreasing in X, because in the Exponential family $\delta_\pi(X)$ is always monotone nondecreasing (see Berger (1986)) and two nondecreasing functions must have nonnegative covariance. e follows from the inequalities:

$$\rho_\pi^2(\theta, \delta_\pi) = 1 - \frac{r(\pi)}{V(\pi)} \geq 1 - \frac{r(\pi_0)}{V(\pi)} \geq 1 - \frac{r(\pi_0)}{V(\pi_0)} = \rho_{\pi_0}^2(\theta, \delta_{\pi_0}).$$

f follows from exactly the same argument and Bickel (1981). g follows on observing that

$$\inf_{F,\pi} \rho_{F,\pi}^2(\theta, \delta) = \frac{\inf_{\mathcal{C}_2} V(\pi)}{\inf_{\mathcal{C}_2} V(\pi) + \sup_{\mathcal{C}_1} \text{var}_F(\delta)}. \qquad \square$$

At the closing of this section now, we will give another general result on Bayesian confidence sets for the mean in the general multidimensional Exponential family that follows as a consequence of Proposition 2.

Proposition 4. Let $\underset{\sim}{X} \sim e^{\underset{\sim}{\theta}'\underset{\sim}{x} - \psi(\underset{\sim}{\theta})}(d\mu)$ and let π be any prior on $\underset{\sim}{\theta}$ with a Lebesgue density g satisfying

$$\frac{||\nabla g(\underset{\sim}{\theta})||}{g(\underset{\sim}{\theta})} \leq B, \tag{3.1}$$

where $0 < B < \infty$ is a specified number. Then with a marginal probability of at least $1-\varepsilon$, the ball $B(\underset{\sim}{x}, r)$ has posterior probability $\geq 1 - \alpha$, where $r = \frac{B + \sqrt{\frac{\nu B}{\varepsilon}}}{\alpha}$, where $\nu = E_\pi ||\underset{\sim}{\theta}||_2$ and ε, α are any given numbers in $(0, 1)$.

Remark. If ε is small, it will be very unlikely that the ball $B(\underset{\sim}{x}, r)$ does not have the required $1-\alpha$ coverage for the $\underset{\sim}{x}$ that was observed. Therefore, as Proposition 4 allows complete bypassing of any computing whatsoever with the posterior distribution and $B(\underset{\sim}{x}, r)$ is a handy posterior confidence set, it could be used as a matter of practicality if ε is small.

Proof of Proposition 4: First note that (3.1) ensures

$$||\delta_\pi(\underset{\sim}{x}) - \underset{\sim}{x}||_2 \leq B, \quad \text{uniformly in } \underset{\sim}{x}, \tag{3.2}$$

and hence

$$||\underset{\sim}{b}_\pi(\underset{\sim}{\theta})||_2 \leq B, \quad \text{uniformly in } \underset{\sim}{\theta}. \tag{3.3}$$

(see, for instance, Brown and Hwang (1982)).

Let $V_\pi(\underset{\sim}{x})$ denote $E||\underset{\sim}{\theta} - \delta_\pi(\underset{\sim}{x})||_2^2|\underset{\sim}{X} = \underset{\sim}{x}$. Then, by Corollary 3,

$$\begin{aligned} &P(V_\pi(\underset{\sim}{X}) > \frac{\nu B}{\varepsilon}) \\ &\leq \frac{r(\pi)}{\nu B/\varepsilon} \\ &\leq \varepsilon, \end{aligned} \tag{3.4}$$

where $P(\cdot)$ denotes marginal probability.

Thus, for any $\underset{\sim}{x}$ such that $V_\pi(\underset{\sim}{x}) \leq \frac{\nu B}{\varepsilon}$,

$$\begin{aligned} &P(||\underset{\sim}{\theta} - \underset{\sim}{x}||_2 > r|\underset{\sim}{X} = \underset{\sim}{x}) \\ &\leq \frac{E(||\underset{\sim}{\theta} - \underset{\sim}{x}||_2|\underset{\sim}{X} = \underset{\sim}{x})}{r} \\ &\leq \frac{E(||\underset{\sim}{\theta} - \delta_\pi(\underset{\sim}{x})||_2|\underset{\sim}{X} = \underset{\sim}{x}) + B}{r} \\ &\leq \frac{\sqrt{V_\pi(\underset{\sim}{x})} + B}{r} \\ &\leq \frac{\sqrt{\frac{\nu B}{\varepsilon}} + B}{r} = \alpha. \end{aligned} \tag{3.5}$$

(3.4) and (3.5) now imply the assertion of Proposition 4. □

4. Examples and Illustrations

In this section, we will give a few examples that further illustrate either an interesting phenomenon or a mathematical fact established earlier.

Example 1. Let $X \sim$ Bin (n, p) and let p have the Beta (α, β) prior with density $\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} I_{0 \leq p \leq 1}$, where $\alpha, \beta > 0$. Let $\{P_j(p)\}_{j \geq 0}$ denote the sequence of Jacobi polynomials on [0,1], which form the system of orthogonal polynomials with respect to the weight function $\omega(p) = p^{\alpha-1}(1-p)^{\beta-1}$. Then, the Bayes estimate for the parametric function $\theta = P_{n+1}(p)$ is

$$\frac{\int_0^1 P_{n+1}(p)p^x(1-p)^{n-x}p^{\alpha-1}(1-p)^{\beta-1}dp}{\int_0^1 p^{x+\alpha-1}(1-p)^{n-x+\beta-1}dp}. \tag{4.1}$$

However, $p^x(1-p)^{n-x}$ is a polynomial of degree n and so equals $\sum_{j=0}^{n} c_j P_j(p)$ for some constants $\{c_j\}$. Hence,

$$\begin{aligned}&\int_0^1 P_{n+1}(p)p^x(1-p)^{n-x}p^{\alpha-1}(1-p)^{\beta-1}dp\\&=\sum_{j=0}^{n} c_j \int P_{n+1}(p)P_j(p)\omega(p)dp\\&=0,\end{aligned}$$

as the system $\{P_j\}$ is orthogonal.

Hence, from (4.1) the Bayes estimate $\delta_\pi \equiv 0$ and so $\rho_\pi(\theta, \delta_\pi) = 0$ although π is not degenerate.

For $n = 1$, θ works out to the interesting function $p(1-p) = \text{var}\ (X)$ if $\alpha = \beta = 1$.

Example 2. Let $X_1, \ldots, X_n$ be iid $N(\theta, 1)$ with the sufficient statistic $\bar{X} \sim N(\theta, \frac{1}{n})$ and let θ have a prior π in the class

$$\mathcal{C} = \{\pi\colon\ E_\pi(\theta) = 0,\ V(\pi) = 1\}, \tag{4.2}$$

Recall from (2.9) that

$$\begin{aligned}&1-\rho_\pi^2(\theta,\delta_\pi)\\&=\frac{r(\pi)}{V(\pi)}\\&=r(\pi)\\&=\frac{1}{n}-\frac{1}{n^2}I(m),\end{aligned} \tag{4.3}$$

by the Brown identity on Bayes risks (Brown (1971)), where $m(x) = \int \sqrt{\frac{n}{2\pi}} e^{-\frac{n}{2}(x-\theta)^2} d\pi(\theta)$ and $I(\cdot)$ denotes the Fisher information functional.

Since the variance of the marginal distribution equals $1 + \frac{1}{n}$ for any π in $\mathcal{C}$, one has

$$\inf_{\pi \in \mathcal{C}} I(m) = \frac{n}{n+1}, \tag{4.4}$$

as normal distributions have the Fisher information minimization property if the variance is fixed; see Huber (1981). From (4.3) and (4.4), therefore,

$$\begin{aligned} &\sup_{\pi \in \mathcal{C}}(1 - \rho_\pi^2(\theta, \delta_\pi)) \\ &= \frac{1}{n} - \frac{1}{n(n+1)} = \frac{1}{n+1} \\ &\Rightarrow \inf_{\pi \in \mathcal{C}} \rho_\pi(\theta, \delta_\pi) = \sqrt{\frac{n}{n+1}}. \end{aligned} \tag{4.5}$$

On the other hand, as the operator $I(\cdot)$ is convex (again see Huber (1981)), $I(m) \leq \int I(N(\theta, \frac{1}{n}))d\pi(\theta) = n$. But, if one considers an element of $\mathcal{C}$ of the form $p\delta\{-c\} + p\delta\{c\} + (1-2p)\delta\{0\}$, where $c = \frac{1}{\sqrt{2p}}$ and $\delta\{\cdot\}$ denotes point mass, then for the corresponding marginal m_p, $I(m_p) \to n$ when $p \to 0$. Together, these imply

$$\begin{aligned} &\inf_{\pi \in \mathcal{C}}(1 - \rho_\pi^2(\theta, \delta_\pi)) \\ &= \frac{1}{n} - \frac{1}{n^2} \cdot \sup_{\pi \in \mathcal{C}} I(m) \\ &= \frac{1}{n} - \frac{n}{n^2} \\ &= 0 \\ &\Rightarrow \sup_{\pi \in \mathcal{C}} \rho_\pi(\theta, \delta_\pi) = 1. \end{aligned} \tag{4.6}$$

(4.5) and (4.6) show that $\rho_\pi(\theta, \delta_\pi)$ converges to 1 as $n \to \infty$ *uniformly* over π in $\mathcal{C}$.

In general, exact minimization or maximization of the Fisher information functional in a class of probability measures is quite difficult; see Bickel and Collins (1983) and Huber (1981).

Example 3. An extremely large literature has now accumulated on robust estimation of a location parameter; a recent reference is Staudte and Sheather (1990). The common criterion is the maximum variance of the asymptotic distribution of a candidate estimate.

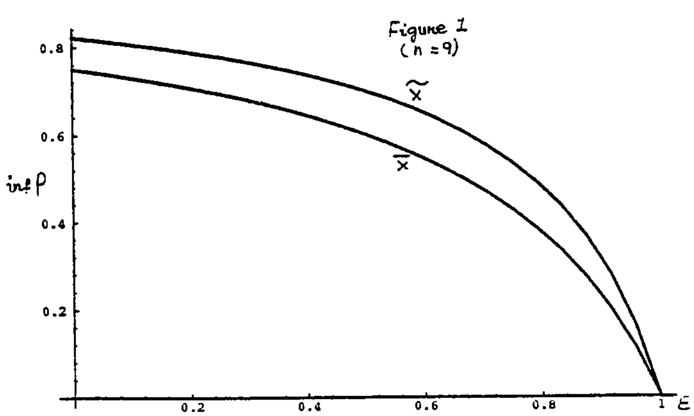
Comparatively far less work has been done on the finite sample variance of popular estimates at various sampling models; one reference is Gastwirth and Cohen (1970). Also see pp. 364 in Lehmann (1983). We will study the correlation of an estimate with a location parameter in finite samples and see how two different estimates compare. The estimates, the priors, and the sampling models are as follows:

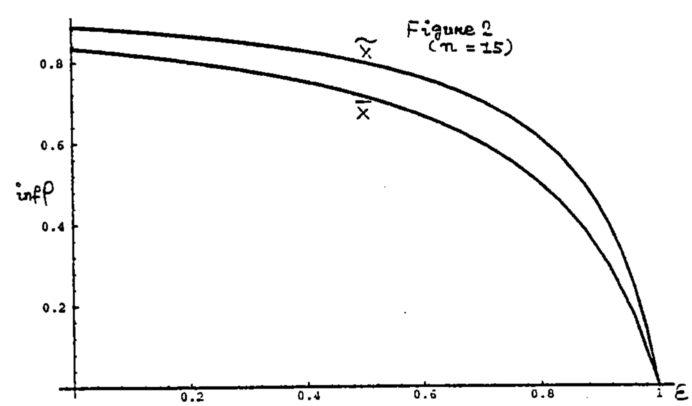
$$\left.\begin{array}{ll} \text{estimates:} & \bar{X} \text{ (mean)}, \tilde{X} \text{ (median)} \\ \text{priors:} & (1-\varepsilon)N(0,1)+\varepsilon Q, \quad Q \text{ symmetric about } 0 \\ & \text{(the Huber class)} \\ \text{sampling models:} & t \text{ distributions with mean } \theta \text{ and } m \text{ degrees} \\ & \text{of freedom, } m \geq 3. \end{array}\right\} \tag{4.7}$$

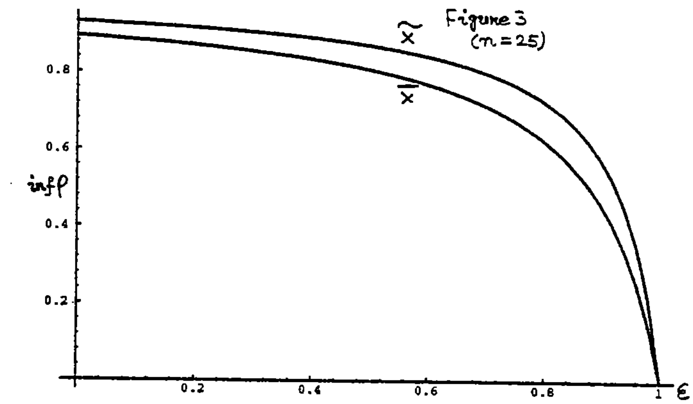
The quantity to be studied is the minimum correlation $\inf_{\pi,f} \rho_{\pi,f}(\theta,\delta)$ in finite samples for each of the two estimates $\bar{X}$, $\tilde{X}$. Recall from (2.4) that

$$\begin{aligned} \inf_{\pi,f} \rho_{\pi,f}(\theta,\delta) &= \frac{\inf V(\pi)}{\inf V(\pi) + \sup\ \mathrm{var}_f(\delta)} \\ &= \frac{1-\varepsilon}{1-\varepsilon+\sup\ \mathrm{var}_f(\delta)}. \end{aligned} \tag{4.8}$$

For $\delta = \bar{X}$ or $\tilde{X}$, exact variance expressions for finite n are available and hence (4.8) is calculable as a function of n and ε. Figure 1 plots (4.8) as a function of ε for $n = 9$, Figure 2 for $n = 15$, and Figure 3 for $n = 25$. The median is always on top, although the excess gradually decreases with n.







Example 4. An interesting questions is the following: suppose δ is a given natural estimate of a parameter θ; which priors give the largest and the smallest value of the correlation $\rho_\pi(\theta, \delta)$? If δ is a Bayes estimate itself with respect to some π_0, often π_0 will maximize $\rho_\pi(\theta, \delta)$. What if δ is not a Bayes estimate?

Let $X \sim$ Bin (n, p) and consider the estimate $\delta(X) = \frac{X}{n}$ for p. Suppose the prior for p is an arbitrary unimodal distribution on $[0, 1]$. We will work out $\sup_\pi \rho_\pi(\theta, \delta)$ and $\inf_\pi \rho_\pi(\theta, \delta)$ and the particular priors at which they are attained.

Denote $E_\pi(\theta)$ by c_1 and $E_\pi(\theta^2)$ by c_2; from (2.4), it follows that

$$\rho_\pi^2(\theta, \delta) = \frac{c_2 - c_1^2}{c_2(1 - \frac{1}{n}) + \frac{c_1}{n} - c_1^2}. \tag{4.9}$$

This is increasing in c_2 for given c_1, by calculus. Now, the moment set

$$\begin{aligned} &\{(c_1, c_2) : \ \pi \text{ is unimodal on } [0, 1]\} \\ &= \{(c_1, c_2) \colon \ 0 \le c_1 \le 1, \ c_1^2 \le c_2 \le \xi(c_1)\}, \end{aligned} \tag{4.10}$$

where $\xi(c_1) = \frac{2}{3}c_1$ if $c_1 \le \frac{1}{2}$ and $\frac{4}{3}c_1 - \frac{1}{3}$ if $c_1 \ge \frac{1}{2}$ (see DasGupta (1995). Figure 4 gives the moment set (4.10).

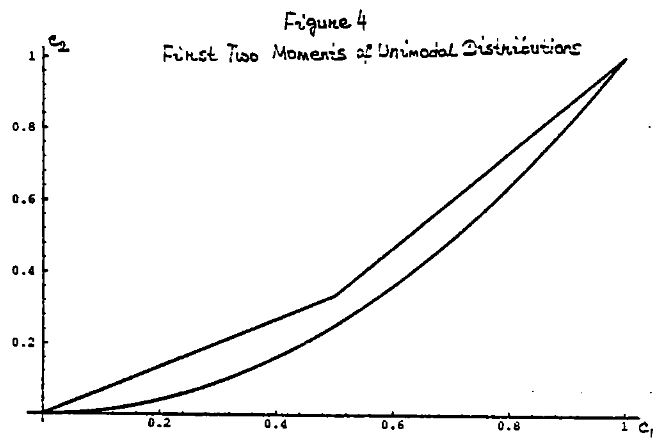


Figure 4
First Two Moments of Unimodal Distributions

From (4.9) and (4.10),

$$\sup_{\pi} \rho_\pi(\theta, \delta) = \max\{ \max_{0 \le c_1 \le \frac{1}{2}} h(c_1), \max_{\frac{1}{2} \le c_1 \le 1} h(1 - c_1)\},$$

where

$$h(c_1) = \frac{2 - 3c_1}{(2 + \frac{1}{n}) - 3c_1}. \tag{4.11}$$

By calculus, h is decreasing on $[0, \frac{1}{2}]$, and thus from (4.11), $\sup_{\pi} \rho_\pi(\theta, \delta) = \frac{2n}{2n+1}$, corresponding to a prior degenerate at 0 or 1. On the other hand, a prior degenerate at any interior point gives $\rho_\pi(\theta, \delta) = 0$ and gives the infimum. Thus, in the class of all unimodal priors on [0,1], $0 \le \rho_\pi(\theta, \frac{X}{n}) \le \frac{2n}{2n+1}$.

Example 5. Finally consider the problem of estimating a positive normal mean; i.e., we have $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, 1)$ where $\theta \ge 0$. This is an interesting problem on its own, but we chose it as an example because the UMVUE and the MLE of θ are not linear functions of one another and so a basis exists for comparing their correlations with θ under various priors for the mean θ. Indeed, the UMVUE is $\bar{X}$ and the MLE is $\max(\bar{X}, 0) = \bar{X}_+$. We will consider gamma priors with density $\frac{e^{-\theta}\theta^{\alpha-1}}{\Gamma(\alpha)}$, $\alpha > 0$, for θ. For $\bar{X}$, we have the easy formula (2.4) and for $\bar{X}_+$, we use the general formula (2.1) in order to calculate correlations. Note that the bias $b(\theta)$ of $\bar{X}_+$ equals

$$b(\theta) = \theta\Phi(\theta\sqrt{n}) + \frac{1}{\sqrt{n}}\phi(\theta\sqrt{n}) - \theta, \tag{4.12}$$

which is used in (2.1) in the numerical computation. Table 1 reports the correlation between θ and each estimate for three values of α and n each.

Table 1: $\rho_\pi(\theta, \delta)$

	n = 1		5		15	
	$\bar{X}$	$\bar{X}_+$	$\bar{X}$	$\bar{X}_+$	$\bar{X}$	$\bar{X}_+$
α						
.5	5774	.6364	.8452	.8795	.9394	.9527
1	.7072	.7404	.9129	.9235	.9682	.9712
2	.8165	.8274	.9535	.9550	.9838	.9840

Thus the MLE consistently enjoys a higher correlation but the excess is negligible for large n or large α.

References

Berger, J. (1986). Statistical Decision Theory and Bayesian Analysis, Springer Verlag, New York.

Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted, *Ann. Stat.*, **9**, 1301–1309.

Bickel, P. J. and Blackwell, D. (1967). A note on Bayes estimates, *Ann. Math. Stat.*, **38**, 1907–1911.

Bickel, P. J. and Collins, J. (1983). Minimizing Fisher information over mixtures of distributions, *Sankhya, Series A*, 1–19.

Brown, L. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems, *Ann. Math. Stat.*, **42**, 855–903.

Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof, Proc. 3rd Purdue Symp., S. S. Gupta and J. Berger(eds.), Academic Press, New York.

DasGupta, A. (1995). Review of optimal Bayes designs, To appear in vol 13, Handbook of Statistics.

Ferguson, T. (1973). Mathematical Statistics, a Decision Theoretic Approach, Academic Press, New York.

Gastwirth, J. and Cohen, M. (1970). Small sample behavior of some robust estimates of location, *Jour. Amer. Stat. Assoc.*, **65**, 946–973.

Huber, P. J. (1981). Robust Statistics, John Wiley, New York.

Lehmann, E. L. (1983). Theory of Point Estimation, John Wiley, New York.

Staudte, R. and Sheather, S. J. (1990). Robust Estimation and Testing, John Wiley, New York.